# Not only base rates are neglected in the Engineer-Lawyer problem: An investigation of reasoners' underutilization of complementarity

Jean Baratgin, Ira Noveck

Not only base rates are neglected in the Engineer-Lawyer problem:
An investigation of reasoners' underutilization of complementarity

Jean Baratgin
Centre de Recherche en Epistémologie Appliquée

Ira A. Noveck
Centre de Recherche en Epistémologie Appliquée
& Université de Grenoble II

RUNNING HEAD: COMPLEMENTARITY & PROBABILISTIC REASONING

Accepted for publication (Memory and Cognition, October 27, 1998)

Correspondence concerning this article may be directed to Jean Baratgin, CREA-Ecole Polytechnique, 1 rue DESCARTES, 75005 Paris, France or to *baratgin@poly.polytechnique.fr*.
This work was partially funded by a grant from DRET (La Direction de la Recherche et de la Technologie, 94/074) in France. The authors would like to thank the members of GRICE (Groupe de Recherche sur l'Inférence et la Compréhension Elémentaire) for their generous comments: Daniel Andler, Francesco Cara, Guy Politzer, Dan Sperber, and Jean-Baptiste van der Henst. We would also like to thank Gerd Gigerenzer, John Wixted, and an anonymous reviewer for their helpful critiques of an earlier version of this manuscript. Special thanks to Dominique Boutet and the students at IUT of Bretigny sur Orge for their participation in the Experiments.

Abstract

The standard Engineer-Lawyer problem (Kahneman & Tversky, 1973) points to reasoners' failure to integrate mentioned base rate information as they arrive at likelihood estimates. Research in this area nevertheless presupposes that reasoners respect complementarity (i.e., participants ensure that competing estimates add up to 100%). A survey of the literature lends doubt to this presupposition. We propose that participants' non-normative performance on the standard problem reflects a reluctance to view the task probabilistically and that normative responses become more prominent as probabilistic aspects of the task do. Three Experiments manipulated two kinds of probabilistic cues and determined the extent to which a) base rates were integrated and b) the complementarity constraint was respected. Experiment 1 presented six versions of an Engineer-Lawyer-type problem (that varied 3 Levels of cue-to-complementarity and 2 base rates). Results showed that base-rate integration increased as cues-to-complementarity did. Experiment 2 confirmed that Gigerenzer, Hell & Blank's (1988) random draw paradigm facilitates base rate integration; a second measure revealed that it also prompts respect for complementarity. Experiment 3 replicated two of our main findings in one procedure while controlling for the potential influence of extraneous task features. Approaches that describe how probabilistic cues might prompt normative responding are discussed.

Not only base rates are neglected in the Engineer-Lawyer problem:
An investigation of reasoners' underutilization of complementarity

When one needs to consider the likelihood of an outcome, base rates (prior probabilities) become critical pieces of information. For instance, if a survey indicates that a candidate for public office is preferred by 80% of a particular voting bloc, it would then be important to know what is the base rate of this bloc with respect to the general voting population. Obviously, the candidate's chances for election would be considerably higher if the bloc's base rate is 70% of the voting population rather than 30%.

Much research in the psychological literature points out that reasoners often neglect base rate information. In this paper, we focus on one well-known task that highlights this neglect among participants -- the Engineer-Lawyer problem (Kahneman & Tversky, 1973). In the original problem, participants are presented five written portraits of people drawn randomly from a population of 70 lawyers and 30 engineers. In one condition, participants are asked to estimate the probability -- for each of the five portraits -- that the person described is one of the 70 lawyers. In another condition, the participants' task is the same except the base rates are reversed (30 lawyers and 70 engineers) and participants are asked to estimate the probability that the person described is one of the 30 lawyers.

Kahneman and Tversky make two claims from their experiments. The first, which is well known in the probabilistic reasoning literature, is that participants generally neglect base rate information and establish their estimates by using the *representativeness* heuristic, i.e. by judging how well the portraits match the stereotype for an engineer or a lawyer (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1982). Supported by findings such as these, heuristic accounts of reasoning have been very influential in the psychological literature (e.g., see Piatelli-Palmarani, 1993; Slovic, Lichtenstein & Fischhoff, 1988; von Winterfeldt & Edwards, 1986).

This claim has led to some controversy. Many argue that other factors, having little to do with a heuristic explanation, may be responsible for base-rate neglect (for a survey, see Girotto, 1994; Koehler, 1996). These factors focus mostly on the presentation of the task. Base rate information is heeded more reliably, e.g., when portraits are presented before the base-rates (Krosnick, Li & Lehman, 1990) or when the experimental context is more conversationally cooperative (Zukier & Pepitone, 1984; Schwarz, Strack, Hilton & Naderer, 1991; for a review, see Hilton, 1995). Thus, many are dubious about the implications from Kahneman & Tversky's studies, most notably Gigerenzer and his colleagues (Gigerenzer and al., 1989; Gigerenzer, 1991, 1993, 1994). In one study exemplary of Gigerenzer's objections, Gigerenzer, Hell and Blank (1988) show how participants' estimates reveal consideration of base rates significantly more often when participants actually draw the portraits from urns whose contents are known to have specific a priori base rates. The aim of the present work is not to take sides in this debate but to better understand what blocks participants from using base-rate information on problems of this sort. This leads us to Kahneman & Tversky's second, lesser-known, claim which is that participants nevertheless comply with the task's complementarity constraint.

When two groups of participants are presented identical base rates (e.g., 30 lawyers and 70 engineers) and one group is asked to estimate the probability that a given portrait is one of the 30 lawyers and the other group that the portrait is one of the 70 engineers, Kahneman and Tversky found that participants' estimates in the two groups to be complementary (estimates from these two groups add up to 100%). This makes sense because laws of probability indicate that *P(Lawyers)=1-P(Engineers).* That is, if participants were to be asked to estimate the likelihood that an engineer-like portrait is one of the 30 engineers, one would find median estimates to be, say, 90% and when asked if he was one of the 70 lawyers, median estimates would be 10%. This indicates that participants respect the complementarity constraint, even as they are ignoring base rates. From that moment forward, Kahneman and Tversky's Engineer-Lawyer tasks requested participants to supply estimates for only one of the occupations (say, only for engineers) because it was assumed that estimate requests in the complementary condition would produce complementary results. Most researchers in the field have followed this practice (what we call an *asymmetric* approach) of presenting only one set of base rates (e.g., Carrol & Siegler, 1977; Wells & Harvey, 1978; Gigerenzer, Hell & Blank, 1988; Schwarz, Strack, Hilton & Naderer, 1991). As we will show, this practice may well be based on a false positive result and prevents one from fully understanding the reasoning processes involved on the Engineer-Lawyer task.

Their complementarity claim is questionable for the following two reasons. First, evidence for complementarity -- when reported -- has not been verified. Swieringa, Gibbins, Larsson & Sweeney (1976) did not replicate the complementarity finding with calculated medians and Kronick, Li & Lehman (1990) failed to replicate with means. Second, there is direct evidence revealing participants' tendency not to comply with the complementarity constraint on the Engineer-Lawyer task when estimates for each of the two hypotheses is requested. In a *symmetric* design, Davidson and Hirtle (1990) required participants to give two likelihood judgments for each portrait: one that determined the likelihood that the person in question was an engineer and another that the person was a lawyer. Before normalizing their data, they found that of 120 pairs of estimates, 24 pairs (20%) added up to amounts other than 100% and that the sums ranged from 90% to 120%. Despite such findings, authors have not drawn any links between their results and the reasoning processes involved; research into this task typically ignores or normalizes this sort of violation. We assume that the failure to respect the complementarity constraint -- what we will call the *neglect to respect complementarity* -- occurs frequently when participants make estimates on the standard Engineer-Lawyer problem and indicates that there is more going on with the task than simply a failure to consider base rates.

That a substantial percentage of participants tend to neglect considerations of complementarity has been investigated in fact (Cohen, Dearnaley and Hansel, 1956; Alberoni, 1962; Marks and Clarkson, 1972; Teigen, 1974, 1983; Einhorn and Hogarth, 1985; van Wallendael and Hastie, 1990; Asare and Wright, 1995). When there are two alternatives, one finds both supra-additivity, where the sum of participants' estimates is over 100%, and sub-additivity, where the sum of their estimates is below 100%. This literature reveals both that the neglect grows with increased task complexity and that it can be eliminated. Although the neglect to respect complementarity does not appear to be as severe as the base rate neglect, it is important to bear in mind that once an experiment tests for complementarity it implicitly cues it. Thus, a tendency to neglect complementarity on Standard versions of the Engineer-lawyer task may well be more common than one is led to believe by findings such as Davidson and Hirtle's.

A normative analysis of the problem shows how an underappreciation of alternativeness can affect participants' estimates on tasks such as the Engineer-Lawyer problem. Bayes' Theorem in (1), which is the standard reference for such problems, indicates that there are two places that require normative reasoners to consider competing hypotheses in approaching the Engineer-Lawyer problem. In the event participants do not consider the alternative hypothesis, the denominator is incomplete causing the estimates to depart from Bayesian norms. Similarly, in the event participants do not consider complementarity, $P(H|D)+P(\text{not } H|D)=1$, their estimates will also vary from those predicted by Bayesian norms.

1. $$P(H|D)=\frac{P(D|H)*P(H)}{P(D|H)*P(H) + P(D|\text{not-}H)*P(\text{not-}H)}$$

As stated earlier, we find it untenable to suppose, as many in the literature do, that participants do not appreciate the alternativeness implicit in base rates but that they do appreciate the alternativeness implicit in the complementarity constraint. This is why the goal of the present work is to first verify that complementarity to 100% tends to be neglected and then to investigate how efforts to counter this neglect can facilitate normative responses on the Engineer-Lawyer problem.

We propose that the consideration of complementarity on the one hand and base rates on the other are part and parcel of a unitary cognitive ability to think probabilistically. Furthermore, we propose that focus on the portraits has the potential to draw away the resources required to reason probabilistically (Baratgin, 1995). The Standard Engineer-Lawyer problem, in our view, leads to a low occurence of base-rate integration because reasoners focus their resources on the portraits and not on the problem’s probabilistic aspects. Therefore, we predict that attention paid to relevant probabilistic notions (whether it be complementarity or base rates) will prompt higher rates of normative responding on such tasks. This interpretation motivates the experiments that follow.

In Experiment 1, we exploit people’s accessibility to probabilistic reasoning by providing two levels of a cue to complementarity on Engineer-Lawyer-type problems. It is proposed that such cues will induce normative responses because they compel participants to take into account probabilistic aspects of the task in general. These two cues differ in degree. Therefore, we expect a strong cue to complementarity to prompt higher rates of normative responding than a less compelling one. In Experiments 2 and 3, we take the converse approach: we investigate a version of the problem that has been previously shown to readily prompt base-rate integration -- the Urn problem (Gigerenzer et al., 1988) -- in order to determine the extent to which it prompts participants to respect complementarity. We anticipate that the Urn problem, which highlights probabilistic features, will prompt relatively few complementarity violations.

## Experiment 1

We adapted the Engineer-Lawyer problem to French participants in order to preserve its high degree of diagnosticity (Fischhoff and Bar-Hillel, 1984). We chose two occupations whose stereotypes are clearer to French sensibilities, that of a mathematics teacher and that of a French literature teacher. This variation of Kahneman & Tversky's Engineer-Lawyer problem is presented to participants under one of three conditions: In the Standard condition, participants are required to estimate the likelihood that a description reflects one hypothesis (math teacher). In the Suggested-Complementarity condition, participants are required to estimate the likelihood that the given description reflects one hypothesis (math teacher) and then the other (French literature teacher). The Induced-Complementarity condition is identical to the Suggested-Complementarity condition except that participants are further reminded that their estimates must add up to 100%.  The Induced-Complementarity condition is expected to yield estimates closest to those anticipated by Bayesian reasoning.

Method

*Participants*

One hundred and twenty first-year students enrolled in a general communication course at the *Institut Universitaire Technique* (IUT) de Bretigny sur Orge (France) took part in the experiment. They had no knowledge about studies on base rate neglect and Bayesian theory. They were tested in small groups.

*Design and Procedure*

This was a 2 (High [70%] vs. Low [30%] base rate of mathematics teachers) X 3 (task instructions:  Standard, Suggested-Complementarity and Induced-Complementarity) between-participants design. Each participant was presented a set of portraits and was assigned randomly to one of the six conditions. There were twenty participants in each condition.

*Material*

Following Kahneman and Tversky (1973), participants in the Low base-rate condition received the following instructions (translated from French):

> The ministry of education conducted a research project concerning the psychological profile of school teachers. A panel of psychologists interviewed and administered personality tests to 100 teachers. There were 30 teachers of mathematics and 70 teachers of French literature. For each teacher, a portrait was written by a psychologist. You will find on your forms five portraits, chosen at random from the 100 available portraits. For each portrait, please indicate the chances (in percent) that the person described is one of the 30 math teachers. You are asked to give your answer on a scale from 0% to 100%.

The High base-rate condition differed only in that participants were told there were 70 math teachers and 30 French literature teachers and that they had to indicate the chances that the person described is one of the 70 math teachers. The remainder of the section will use the Low base-rate condition as an example.

This kind of introductory instruction was followed by five psychological portraits that had been previously tested in a pilot study. Two were designed to have a strong informative message and thus to be highly diagnosing. One implied that the person described (Jacques) is a math teacher while a second (Anne) indicated that she is a French literature teacher. A third (Françoise) was designed to be less diagnostic yet suggestive of a French literature teacher. Her portrait was drawn from a real portrait (in Gigerenzer and al., 1988). A fourth was uninformative (Paul). Finally, Raphaël's portrait was informative but inconsistent (as in Ginosar and Trope, 1980). Questionnaires concerning the five portraits were prepared in one of five different random orders. Below are the five portraits presented:

> Jacques is 45 years old. He is married and has four children. He shows modest interest in politics and current affairs. He spends most of his spare time doing carpentry, crosswords and yachting.
>
> Anne is 28 years old. She is single, cheerful, and enthusiastic about her job. She has a special interest in fine arts. She runs a small gallery in her spare time where she exhibits the work of young artists. She also likes the seaside and to take walks.
>
> Françoise is 48 years old and has remained single. She is down to earth, earnest and seems to be appreciated by her pupils. She is very politically engaged and is active in a trade union. Her main hobbies are mountain hiking and travelling to distant countries.
>
> Paul is 50 years old. He is married and has two children. He is a man of a great intellectual capacity and is very motivated. He is very successful at his job. He is appreciated by his colleagues.
>
> Raphaël is 35 years old. He is separated from his wife and is raising his daughter by himself. He is rather outgoing and likes jokes and plays on words. He enjoys going to the theater and visiting museums. He is a good chess player, as well. He has started a chess club in his school.

Participants were required to estimate for each portrait the chances (in percent) that the person described is one of the mathematics teacher. For example, with respect to Anne in the Low base-rate condition, a participant's estimate was requested in the following manner:

> The chances (in percent) that Anne is one of the 30 teachers of mathematics is __%

In the Suggested Complementarity/Low base-rate condition, the procedure was identical except that the second to last sentence of the introductory paragraph ended with "...and the chances (in percent) that the person described is one of the 70 teachers of French literature" and the task instruction read as follows (e.g., for Anne):

> The chances (in percent) that Anne is one of the 30 teachers of mathematics is ___%.
>
> The chances (in percent) that Anne is one of the 70 teachers of French literature is ____%.

The Induced Complementarity condition was identical to the Suggested-Complementarity condition except that participants were required to provide probabilities that add up to 100% :

> Please note that the sum of the two percentages must be 100%. For example, Anne is either a mathematics teacher or a French literature teacher.

Finally, all participants were asked to assess the confidence they had in their answers. They were told that some experts (different from the psychologists who wrote the portraits) had been given the same exercise with the same five portraits. Participants had to rate on a scale, ranging from 0% to 100%, the chances that they had given the same answer as the experts for each portrait and more generally over the entire task. They were entitled to look back in order to revisit their original estimates but were not allowed to change them. Most participants, however, provided their confidence ratings without looking back.

*Results*

We follow Wells and Harvey (1978) and Gigerenzer et al. (1988) by applying a set of fine-grained measures based on participants actual performance and by determining how the same participants would have performed under idealized, Bayesian conditions. Although descriptive statistics often prove to be sufficiently powerful, we provide a few critical statistical tests to back up the claims and to provide appropriate landmarks. We then analyze participants' performance with respect to complementarity violations (where applicable) and with respect to provided confidence judgments. A summary of the results is presented in Table 1.

---------------------------------------------

Insert Table 1 about here

---------------------------------------------

In order to determine the extent to which participants use base rate information, we assess the degree to which the two associated groups (High/Low base rates) in a given task-instruction condition yield different scores for each portrait. In the event that participants do not use base rates, the two groups' estimates ought to be similar[1]. The first column of Table 1 provides the mean probabilities that the person described is a mathematics teacher given the Low base rate condition. The second column presents the mean probabilities in the High base rate condition. For each portrait, we compare the difference between the mean based on the percentages given by participants in the High base rate group and the one worked out for the Low base rate group. We call this the ***D****ifference between the* ***M****eans of the* ***H****igh and* ***L****ow Base Rate Groups* or *DMHL*. DMHL appears in the third column.

In the Standard condition, the overall DMHL (across the five portraits) was 6.8%. This difference indicates that our findings are in line with results from previous studies; Kahneman and Tversky (1973) found a DMHL of 5% and Gigerenzer, Hell and Blank 9.8% (1988). In the Suggested-Complementarity condition, the difference averaged across the five portraits was substantially higher (DMHL=11.2%) than in the Standard condition. This indicates that the moderate cue to consider both hypotheses led a greater percentage of participants to include base rate information. We analyzed separately the results of those participants who had complied with the complementarity constraint in the Suggested-Complementarity condition. This is presented on Table 2. This subset of participants (n=29) revealed a DMHL of 16.4% and the difference between the High and Low groups -- for each of the five portraits -- is sizeable. These results are remarkably similar to those found in the Induced-Complementarity condition (DMHL=17.5%), which led to the biggest overall difference between the High and Low base rates. Clearly, base rates are integrated more as cues to complementarity become more acute.

---------------------------------------------

Insert Table 2 about here

---------------------------------------------

We now carry out a traditional analysis (tests of significance) on the DMHL data. For the Standard condition, the mean DMHL of 6.8% is significantly different from predictions based on chance, $t(38)=1.73$, $p<0.05$, indicating that our participants generally did use base rates. However, none of the portraits, treated separately, yielded a DMHL that was significant at the .05 level. Two (Anne, and Jacques) yielded differences that were marginally significant (t-tests yielding $p < .1$) and the other three (Françoise, Paul and Raphaël) gave results that were more decidedly not significant. In the Suggested-Complementarity condition, the average DMHL proved highly significant, $t(38)=2.84$, $p <0.005$). Two portraits yielded significant differences ($t(38)=2.2$, $p <0.025$ for Anne and $t(38)=1.79$, $p <0.05$ for Jacques) and the other three were marginally significant. We then analyzed the results of the subset of participants (n=29) who had complied with the complementarity constraint in the Suggested-Complementarity condition: The overall DMHL of 16.4% was significant ($t(27)=5.16$, $p <0.0005$) and significant differences were found between the High and Low groups for each of the five portraits ($t(27)=2.42$, $p<0.01$ for Anne, $t(27)=1.83$, $p<0.05$ for Françoise, $t(27)=2.11$, $p<0.025$ for Jacques, $t(27)=2.65$, $p<0.01$ for Paul and $t(27)=1.93$, $p<0.05$ for Raphaël). The Induced-Complementarity condition led to the biggest overall difference between the High and Low base rate groups (DMHL=17.5%, $t(38)=4.58$ $p <0.0005$). Each of the five portraits therein revealed significant differences with respect to DMHL: $t(38)=2.67$, $p<0.01$ for Anne, $t(38)=2.55$, $p<0.01$ for Françoise, $t(38)=2.97$, $p<0.005$ for Jacques, $t(38)=2.34$, $p<0.025$ for Paul and $t(38)=2.09$, $p<0.025$ for Raphaël.

To verify the above findings across conditions, we carried out a One-way Between-Subject ANOVA in which *Degree of Cues to Complementarity* was the independent variable and the DMHL for each of the portraits was treated as a dependent measure. The result was significant, $F(2, 12)=34.038$, $p < .001$. Followup Scheffé tests showed significant differences both between the Standard and Suggested-Complementarity conditions ($p < .05$) and between the Suggested- and Induced-Complementarity conditions ($p <.005$). Lastly, one finds a significant difference when comparing the DMHL of those (29) who respect complementarity in the Suggested-complementarity set and those (11) who do not, $t(8)=2.20$, $p < 0.05$.

We now analyze the differences between the estimates provided by our participants and idealized Bayesian estimates worked out from equation 1. We followed Wells and Harvey's approach (1978) and calculated -- for each participant -- the probability estimate a participant would have given if she were in the other base-rate condition. For example, the estimates given by a participant in the Low base-rate condition can be transformed to determine what her responses would have been if she were assigned to the High base-rate group. Conversely the same can be done in the other group. For statistical assessment, the means of the Bayesian-derived estimates for each portrait were then compared against the mean probability estimates of the actual group. We call this difference the *Bayesian Estimate minus the Actual Estimate* (BE-AE). To distinguish the BE-AE for the Low base-rate condition from that in the High base-rate condition, we call the former BE-$AE_{Low}$ and the latter BE-$AE_{High}$. BE-$AE_{Low}$ and BE-$AE_{High}$ are presented in the fourth and fifth columns of Table 1. The sixth column of Table 1, called BE-$AE_{Mean}$ calculates the mean of these two differences. If the mean of the differences is null or low, that would indicate normative responding. In contrast, a relatively high mean of the differences indicates non-Bayesian inference-making.

To be consistent with Gigerenzer et al. (1988), we include BE-$AE_{Low}$. We also include an index developed from one that they dubbed *b* and corresponds (in our terms) to:

(2) $$\frac{\text{DMHL}}{\text{DMHL} + \text{BE-AE}_{Low}}$$

We point out, however, that we do not suppose that the two groups (High and Low) behave symmetrically (Baratgin & Andler, in preparation). Consequently, we call (2) above $b_{Low}$ and construct a more complete index which we call $b_{Mean}$:

(3) $$b_{Mean} = 100*1/2(b_{Low} + b_{High}) \quad \text{or}$$

$$b_{Mean} = 100*1/2(\text{DMHL}/(\text{DMHL} + \text{BE-AE}_{Low}) + \text{DMHL}/(\text{DMHL} + \text{BE-AE}_{High}))$$

This index varies from 0 (no indication of base rate utilization) to 100 (perfect Bayesian treatment of base rate information). The indices $b_{Low}$ , $b_{High}$, and $b_{Mean}$ are presented in the last three columns of Table 1.

Although it is obvious that participants generally do not approach these problems in a thoroughly normative manner, one notices that participants' estimates approach Bayesian-derived probabilities as cues to complementarity increase. The overall BE-$AE_{Mean}$ starts out high at 22.1% in the Standard condition and decreases to 17.6% in the Suggested-Complementarity condition and to 12.4% in the Induced-Complementarity condition. The same holds for BE-$AE_{Low}$ which steadily declines from 21%, to 16.5% and, finally, to 12.1% for the Standard, Suggested-complementarity, and Induced-complementarity conditions respectively. The index $b_{Mean}$ also reveals increasing Bayesian behavior. In the Standard, Suggested-complementarity, and Induced-complementarity conditions, the index $b_{Mean}$ provided outcomes of 24, 39, and 58, respectively.

We now take a closer look at the Suggested-Complementarity condition in order to see what kinds of errors participants made. Table 3 shows that 11 participants (27.5%) violate the complementarity constraint 34 times overall (55% of error-prone participants were exclusively subadditive, 9% were exclusively superadditive, and the remaining 34% made errors in both directions). This is in line with the findings of Davidson and Hirtle (1974). Our analysis also shows that errors were evenly distributed across the five portraits.

----------------------------------------------

Insert Table 3 about here

----------------------------------------------

In an effort to gain more insight into participants' responding, we considered two possible accounts of these errors. They both stem from the idea that error-prone participants neglect to properly transform their initial estimates (and that this is the step that those in the Induced condition are compelled to carry out); the two proposals reflect two kinds of unrealized transformations. Furthermore, it is plausible that such neglect ultimately affects the indices of base-rate integration. Perhaps once this transformation is completed, the indices would reveal greater base-rate integration. Thus, we carried out these transformations on non-complementary estimates in order to determine their respective effects on the indices. One potentially unrealized transformation is normalization, wherein initial estimates of, say, 60% and 20% retain their respective weights and become 75% and 25%. When we applied this transformation we found that the condition's indices of base-rate integration are not dramatically affected: DMHL = 12, BE-$AE_{Mean}$ =15.6, and $b_{Mean}$ =41. (Also, the reported statistical difference between the Suggested- and Induced-complementarity conditions remains significant). The second, potentially unrealized transformation is "splitting the difference," wherein initial estimates would be converted additively and equally so that the difference from 100% would be distributed equally (in this case, 60% and 20% becomes 70% and 30%, respectively). When errors are transformed by hand in this manner, the condition's indices of base-rate integration are again not noticeably different: DMHL = 11, BE-$AE_{Mean}$ =19, and $b_{Mean}$ =37.9. Thus, we conclude that a failure to carry out a (singular kind of) transformation after arriving at initial estimates does not affect the base-rate integration indices in the Suggested-complementarity condition. The significance of these results will be taken up in the Discussion.

Finally, we analyze participants' confidence measures and point to two findings. First, we find no difference in confidence judgements with respect to the High and Low conditions.

This is reassuring because it shows that there is nothing intrinsic to the relative weight of the base rate information. Second, we find that confidence grows systematically as cues to complementarity increase. Participants report higher confidence as the task's complementarity aspect becomes more compelling: The Standard, Suggested-Complementarity, Induced-Complementarity conditions yielded mean confidence ratings of 43.2%, 53.4%, and 61.5% respectively. Highly diagnosing portraits, Anne and Jacques, tend to prompt higher confidence estimates (56.3% and 54.8%, respectively) than the intermediate ones, Paul and Raphaël (52.4% and 52.2%, respectively). The lowest level of confidence was associated with Françoise (47.6%) who provided the mixed message.

*Discussion*

Evidently, normative responding is facilitated when a task points to the two alternatives and is further facilitated when it compels participants to respect the complementarity constraint. We point to our three main findings. First, participants were more likely to employ base rates in both the Suggested- and Induced-complementarity conditions than in the Standard condition. This increase is evident for each of the five portraits.

Second, there is a difference between the Induced- and Suggested-complementarity conditions. This is especially important because it reveals a facilitative role for the cue when it is more strictly related to complementarity. There is no conceivable way to induce complementarity without first requesting two estimates, as was done in the Suggested-complementarity condition. Thus, we suppose that the additional cue in the Induced-complementarity condition (to make sure that the two estimates add up to 100%) served to promote probabilistic reasoning and, in turn, the integration of base rates. Note that this was accomplished without cues to the base rates. An alternative hypothesis is that the difference between participants in the Suggested- and Induced-complementarity conditions is an artifact due to participants' compliance with the instructions. That is, a participant in the Induced-complementarity condition, after first generating intuitive estimates that are not complementary, applies a transformation that makes the estimates sum to 100%. According to this hypothesis, it is the transformation that accounts for the changes in the indices and not the cue. Our analyses of errors ruled out this alternative. When the complementarity violations of the Suggested-complementarity condition were normalized (or adjusted by "splitting the difference"), indices in that condition remained low relative to those in the Induced-complementarity condition.

Third, confidence level increases as complementarity is emphasized. This indicates that an appreciation of probabilistic aspects of the task is important to participants' confidence. There may be another interpretation of the increasing confidence measures across task conditions: higher confidence measures may be a byproduct of increasing task-demands. More required effort may be giving participants a more confident impression. Nevertheless, it is clear that participants are increasingly more successful and confident as they are encouraged to integrate probabilistic information.

## Experiment 2

We examine whether Gigerenzer et al.'s Urn problem (1988), a version of the task which highlights its probabilistic features and typically demonstrates integration of base rate information, provides cues to probabilistic reasoning that latently engage respect for complementarity. If our hypothesis is correct -- that normative performance on the task is linked to probabilistic reasoning viewed more generally -- we should find that Gigerenzer et al.'s experimental paradigm leads to low rates of complementarity violation. However, if our indices show base rate integration while producing high rates of complementarity violations, it would concur with the assumption made by many investigators -- that complementarity and base-rate integration are independent processes.

In the experiment that follows, we present Gigerenzer et al.'s Urn task and assign subjects to the High and Low base rate conditions with instructions to determine estimates with respect to each of the two professional groups. The two estimates serve as a dependent measure of respect for complementarity.

### Method

*Participants*

Forty students, who came from the same population of students as those in Experiment 1, took part in this study.

*Design and Procedure*

Participants were randomly assigned to one of two conditions (Low and a High base rates of math teachers). The task modelled the instructions after those in the Suggested-Complementarity condition of Experiment 1 but implanted a random draw like the one carried out by Gigerenzer et al. (1988).

*Material*

Each participant was individually interviewed. He (or she) sat facing the interviewer who was seated at a desk. The first four sentences of the instructions were identical to those found in the Suggested-complementarity condition of Experiment 1. The rest is presented in the excerpt below, which comes from the Low base rate condition:

> Ten (10) of these portraits are presented in front of you. Among them you will find three portraits of math teachers indicated with an M and the seven remaining ones are portraits of teachers of French literature; they are marked with an F. In the five urns in front of you there are copies of the same ten portraits. You are asked to draw at random one portrait from each urn. For each portrait, please indicate the chances (in percent) that the person described is one of the 30 math teachers and the chances (in percent) that the person described is one of the 70 teachers of French literature. You are asked to give your answer on a scale from 0% to 100%.

The participants gave their two estimates for each portrait after each draw. The order of presentation of the urns was systematically shuffled to prevent a spurious order effect. We ensured that we had ten copies of the same portrait in each urn while we let our participants believe they were drawing at random. Participants were entitled to ask any questions until they began drawing from the first urn. After that, questioning was forbidden. The High base-rate condition, of course, inverted the base rates provided in the above instruction.

*Results*

We refer to the present results as outcomes of the Urn condition and we analyze participants' responses in a manner similar to the one used in Experiment 1. The mean difference between the High and Low groups (DMHL) across the five portraits in the Suggested-Complementarity Urn condition was well above zero (DMHL=14.7), indicating that participants largely processed base rates. Other indices also point to Bayesian responding (BE-AE$_{Mean}$ = 12.87% and b$_{Mean}$ =53). These findings largely confirm those found by Gigerenzer et al. Highly diagnosing portraits (Anne and Jacques) again yielded results closest to Bayesian estimates. The results are presented in Table 4.

-----------------------------------------

Insert Table 4 about here

-----------------------------------------

Only one out of forty violated the complementarity constraint in the Urn condition. This participant violated the constraint on all five estimates, belonged to the High base rate group, and erred in the direction of sub-additivity. Once this participant is removed from the analysis, DMHL increases to 15.77% and BE-$AE_{Mean}$ decreases to 11.88 ($b_{Mean}$ =57). One can see how this one participant can affect, albeit minimally, the indices of base-rate integration. Interestingly, eight participants asked the interviewer whether the sum of the probabilities was required to be 100% during experimental administration (to which the interviewer remained mum).

Although statistical comparisons between the Urn task, on the one hand, and the Suggested- and Induced-complementarity versions of Experiment 1 on the other are not advisable due to procedural differences, a preliminary comparison reveals some noteworthy findings. With respect to DMHL and complementarity, participants' responses in the Urn condition were comparable to those in the Induced-complementarity condition of Experiment 1 and were more normative than those in the Suggested-Complementarity condition. The same holds for the confidence ratings (mean of 60.22% in the present Experiment).

*Discussion*

It is apparent that Gigerenzer et al.'s (1988) Urn problem increases not only the integration of base rate information but participants' respect for the complementarity constraint as well. Thus, our hypothesis is supported: As a task appears more probabilistic in nature, consideration of relevant probabilistic principles increases. When one counteracts the neglect of base rate information on the Engineer-Lawyer problem, one encourages respect for complementarity as well.

Although our results from the Urn condition are largely similar to those found in Gigerenzer et al. (1988), their participants gave estimates that were, on the whole, closer to those predicted by the Bayesian norm. For example, $b_{Low}$ in Experiment 2 provided a mean of 55 whereas the equivalent index is their study produced a mean of 65. This variation may stem from two factors. One factor may be related to the portraits themselves. Gigerenzer et al. portrayed engineers and lawyers and we two kinds of teachers. The other factor may be that Gigerenzer et al. (1988) presented six portraits and we presented five. With six portraits, base rates can be applied more intuitively; e.g., one can conclude that, with 6 portraits and a distribution of 70% engineers and 30% lawyers, a distribution of 4 engineers and 2 lawyers is likely. With five portraits (and the same 70%/30% distribution) two configurations are plausible -- 3 engineers and 2 lawyers or 4 engineers and 1 lawyer. Thus, six portraits may increase Bayesian responding by allowing those participants who appreciate base rate information to constrain the problem (Baratgin, 1995).

How can one explain the increased respect for complementarity in the Urn problem? Some might argue that the frequentist presentation of the contents of the urn (showing that seven elements come from one profession and that three come from the other) aids human beings in estimating probabilities normatively (see Gigerenzer, 1993; Gigerenzer and Hoffrage, 1995; Jones, Jones and Frich, 1995; Cosmides and Tooby, 1996). We agree that the Urn version prompts respect for complementarity, but not necessarily for the reasons offered by frequentists. We see two limitations to the frequentist proposal. First, a frequentist presentation appears sufficient but not necessary for increased normative responding. Experiment 1's Induced-complementarity condition prompted estimates indicating base-rate integration at levels comparable to those found on the Urn problem without a frequency format. Second, the presentation of random sampling does not explain what mechanism is triggered (see Griffin and Dukershire (1992) for a critique of Gigerenzer et al.'s (1988) argument). It is proposed here that the random sampling in the Urn condition draws attention to base rates and that this increases base rate use as well as complementarity compliance.

## Experiment 3

The presentation of the Urn in Experiment 2 encourages participants both to integrate base rates and to respect complementarity, but it does not directly present a Control condition. If one were to treat the Suggested-Complementarity condition of Experiment 1 as a Control (because it offers the same task instructions but lacks the Urn presentation), results indicate that the Urn context provides strong cues to normative responding (see Tables 1 and 4). In the present Experiment, we determine whether this observation is replicable in one overarching procedure that licenses empirical claims.

In fact, there are two features of the Urn condition of Experiment 2 that distinguish it from the Suggested-complementarity condition of Experiment 1. One is the presence or absence of the random sampling (Urn) format and the other is the experimenter-participant interaction (the Suggested-complementarity task of Experiment 1 was presented to participants collectively and the Urn task of Experiment 2 individually). Thus, the present Experiment investigates the import of these two task features as we compare performance on the Urn task and the Suggested-complementarity task, which we will call here the no-Urn condition. We anticipate that the presence of the Urn alone will be critical to increasing normative responding as determined by base rate integration and by rates of complementarity violation.

### Method

*Participants*

One hundred and twenty eight students, who came from the same population of students as those in the earlier Experiments, took part in this study.

*Design and Procedure*

As in the earlier experiments, participants were randomly assigned to either Low or High base rates of math teachers. The design was identical to the earlier Experiments, except that two variables were manipulated: The *presence* or *absence* of the Urn format and the *collective* or *individual* manner of experimental presentation. Thus, this was a 2 (High vs. Low base rate of Mathematics teachers) X 2 (Urn vs. no-Urn) X 2 (individual vs. collective presentation) between-subjects design. There were sixteen participants in each condition.

*Materials*

The only novel tasks (with respect to conditions that have not been described earlier) are a) the Suggested-complementarity (*no-Urn*) version presented *individually* and b) the *collective* presentation of the *Urn* condition. The *individual, no-Urn* task simply presented each participant with the problem set of the Suggested-complementarity condition (of Experiment 1) singly in the presence of the experimenter. The *collective, Urn* problem employed the materials of Experiment 2 to groups of eight persons; one member of each group publicly drew out the portraits from the Urn. As in Experiment 2, efforts were made to give the impression that the draws from the five Urns were random. Each urn contained ten (albeit identical) portraits and ten piles of portraits were prepared and displayed so that, once the Experimenter publicly ascertained which portrait was drawn, he could distribute copies to each participant containing the same task instructions as those in the individual condition. (The *collective, no-Urn* task is essentially a replication of the Suggested-complementarity condition of Experiment 1 and the *individual, Urn* task is essentially a replication of Experiment 2.) Given the limited number of groups, one of two random orders of presentation were used throughout the Experiment.

*Results and Discussion*

We analyze participants' responses with respect to the Urn and no-Urn problems as presented individually and collectively. Not surprisingly, the mean difference between the High and Low groups (DMHL) across the five portraits in each of the four conditions was well above zero. The other indices also point to Bayesian responding. These results, as presented in Table 5, largely confirm those found in the two earlier experiments.

----------------------------------------

Insert Table 5 about here

-----------------------------------------

A 2 (Format: Urn vs. no-Urn) X 2 (Task-administration: Collective vs. Individual) Between-subjects ANOVA was carried out with the DMHL serving as a dependent measure. The ANOVA showed a main effect for Format, $F(1, 16)=56.365$, $p<.001$, a marginal, non-significant effect for Task-administration, $F(1,16)=4.12$, $p=.06$, and no significant interaction, $p=.78$. Apparently, collective presentation slightly (and non-significantly) facilitates normative performance. Similar analyses with other dependent measures ($BE\text{-}AE_{Mean}$ and $b_{Mean}$) confirm a main effect for Format, and non-significance for Task-administration, across the Experiment. Thus, one can confidently conclude that the Urn context significantly facilitates normative responding.

Rates of complementarity violations are of particular interest. There were no violations of complementarity in the two Urn conditions and fifteen (23%) such violations across the two no-Urn conditions. The latter is comparable to the rate of such errors in the Suggested-complementarity condition of Experiment 1 (and to the rate found by Davidson & Hirtle, 1990). Eight such errors occurred in the Collective condition and seven in the Individual condition. As before, errors were distributed across portraits. There were no such errors in the two Urn conditions.

In sum, Experiment 3 confirms the findings of the two earlier ones. The Urn condition prompts high rates of base rate integration and no complementary errors. In contrast, the (control) no-Urn condition (which is equivalent to the Suggested-complementarity condition of Experiment 1) is less likely to prompt base rate integration and respect for complementarity. Task administration did not lead to robust effects. Although collective administration tended to influence normative reasoning (according to *DMHL*), this factor did not affect the rate of complementarity errors nor the other, finer measures of base-rate integration.

These findings support two points made earlier. On the one hand, all four of these conditions request two estimates which, according to our hypothesis, prompts some degree of normative reasoning compared to the Standard problem. On the other hand, we do not find thoroughly normative responding in any of the conditions of these Experiments. Even in those problems that strongly underline probabilistic considerations, participants appear to aim for an integration of the protraits with probabilistic information.

## General Discussion

The literature offers two kinds of explanation concerning the neglect to respect complementarity. One explanation is that participants provide estimates of a given hypothesis independently of the alternative hypothesis; i.e., participants adopt a "non-distribution conception of probability" (Teigen, 1983; see also Van Wallendael and Hastie, 1990). This suggests that reasoners lack the competence to consider this fundamental probabilistic principle. Another class of explanations can be drawn out from work on Bayesian reasoning. Many researchers (e.g., Braine, Connell, Freitag, & O'Brien, 1990; Macchi, 1994, 1995) have proposed that participants, in considering base rates, become confused because they invert the given conditional probability and thus make the structure of the problem ambiguous; that is, the estimate *P(not-H½D)* is mistaken for *P(D½not-H).* It is plausible to suppose that this kind of error occurs when complementary estimates are requested as well. This kind of explanation suggests that difficulties arise while performing the task but that competence is still available.

We assume that the competence is available and have argued that facilitative Engineer-Lawyer problems prompt higher rates of correct responses because they draw participants' attention to sources relevant to normative responding. Whereas the original Engineer-Lawyer problem shows how participants may not mechanically draw out probabilistic information, reasoners are clearly capable of applying normative solutions once such information is made relevant to them. More effort is applied to probabilistic information as it becomes more prominent. Cues to complementarity (Experiment 1) is one way to point to the importance of probabilistic reasoning and the Urn context (Experiments 2 and 3) is another. Our proposal (which was motivated by Sperber &Wilson's [1986/1995] Relevance theory) describes why the neglect to respect complementarity is linked to base rate neglect. Participants' *general* failure to accord effort to the probabilistic information in the standard Engineer-Lawyer problem leads to low levels of normative performance.

Although base rates become increasingly important for estimates concerning all the portraits as conditions increase their cues to complementarity, the highly diagnosable portraits (Anne and Jacques) were consistently the most likely to prompt differences between High and Low base rate groups and to prompt estimates revealing normative behavior. This lends doubt to the claim that participants rely exclusively on the *representativeness* heuristic because these informative (stereotyped) portraits provoked the highest rates of normative responses. It appears then that participants seek out information that is coherent and relevant, particularly with respect to two factors: a) the clarity of information in the personality portraits and b) the probabilistic features of the problem.

This work adds to mounting evidence that shows that reasoning, viewed from the point of view of discourse, can prompt normative responses under conditions that make certain aspects of a task relevant (see Sperber, Cara & Girotto, 1995; Noveck, 1997). Work with respect to tasks developed by Kahneman and Tversky have been especially illustrative in this regard. Normative performance on the Linda problem increases when its conversational peculiarities are avoided (Politzer & Noveck, 1991; Hertwig & Gigerenzer, 1996) and normative performance on the Taxi-cab problem increases with the useful presentation of information and carefully worded questions (Macchi, 1995). If one makes clear to participants the speaker's (the experimenter's) communicative intent, normative performance is more likely because the participant has a clearer idea of what the experimenter wants and participants, being relatively good listeners, try to oblige to the best of their abilities. Viewed in this light, one can see that the rate of normative responding varies as a function of the communicative intent in a problem's protocol.

To summarize, we argue that the neglect to respect complementarity is, not only common on the Standard Engineer-Lawyer problem but, closely linked to the well known base rate neglect. We were thus motivated to demonstrate that the two neglects rise and fall together. We showed that two levels of cues to complementarity prompt two corresponding levels of base rate integration and that the Urn task
-- a version known for facilitating base rate integration -- prompts participants to respect complementarity. Based on the nature of the facilitative problems, we take our findings to mean that normative performance on tasks such as these is more likely when a participant is invited to perceive a problem probabilistically.

References

Alberoni, F. (1962), Contribution to the study of subjective probability I", *The Journal of General Psychology, 66,* 241-264.

Asare, S. K. and Wright, A. (1995). Normative and substantive expertise in multiple hypotheses evaluation. *Organizational Behavior and Human Decision Processes, 3,* 171-184.

Baratgin, J. (1995). Sophisme du taux de base et conditions de pertinence. *Ms*.

Baratgin, J. and Andler, D. (In preparation). Les deux apparentes symétrie du problème "des ingénieurs et des avocats".

Beyth-Marom, R. and Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology, 45*, 1185-1195.

Braine, M. D. S., Connell, J., Freitag, J. and O'Brien, D. P. (1990). Is the base rate fallacy an instance of asserting the consequent*?. In: Lines of thinking: Reflections On The Psychology of Thought*, ed. K. L Gilhooly, M. T. G. Leane, R. H. Logie, and G. Erdos. John Wiley and Sons, Vol. 1, pp. 165-180.

Cohen, J., Dearnaley E. J. and Hansel, C. E. (1956). The addition of subjective probabilities. *Acta Psychologica*, *12*, 371-380.

Cosmides, L and Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition, 58*, p. 1-73.

Davidson, D. and Hirtle, S. C. (1990). Effects of nondiscrepant and discrepant information on the use of base rates. *American Journal of Psychology, 103*, 343-357.

Doherty, M. E., Mynatt, C.R. and Tweney, R. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 43*, 11-121.

Einhorn, H. J. and Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 4*, 433-461.

Fischhoff B. and Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory and Cognition*, *12*, 402-410.

Gigerenzer, G. (1991). How to make cognitive illusions disappear. Beyond "heuristics and biases". *European Review of Social Psychology, 2*, 83-115.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In: *Rationality*, ed. Manktelov, K.I. and Over, D.E. Routledge, Londres, 284-313.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In: *Subjective Probability*, ed G Wright and P. Ayton. New York: Wiley, 129-162.

Gigerenzer, G. (1996). On narrow norm and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103, No. 3,* 592-596.

Gigerenzer, G., Hell, W. and Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513-525.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102* , 684-704.

Gigerenzer, G., Swijting, Z., Porter, T., Daston, L., Beatty, J. and Kruger, L. (1989). *The empire of Chance: How probability Changed Science and Everiday Life*. Cambridge University Press, Cambridge.

Ginossar, Z. and Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology, 16*, 228-242.

Girotto, V. (1994). *Il Ragiomento*. Il Mulino, Bologna.

Griffin, D. and Dukeshire, S. (1992). On the use and neglect of base rates: A test of the visual random sampling hypothesis. Department of Psychology, University of Waterloo.

Hertwig, R. & Gigerenzer, G. (1996). The "Conjunction Fallacy" revisited: How intelligent errors look like reasoning errors. *Ms.*

Kahneman, D. (1991). Judgment and decision making: A personal view. *Psychological Science, 2*, 142-145.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-453.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, *103, No. 3*, 582-591.

Koehler, J.J. (1993). The base rate fallacy myth. *Psycoloquy, 4(49)*, base-rate.1.koehler.

Koehler, J.J. (1994). Base rates and the "Illusion" illusion. *Psycoloquy,* 5(9), base-rate.9.koehler

Koehler, J.J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Science*, *19*, 1-53.

Krosnick, J. A., Li, F. And Lehman, D.R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology, 39*, 1140-1152.

Macchi, L. (1994). On the communication and comprehension of probabilistic information. *Psycoloquy,* 5(11*)*, base-rate.11.macchi.

Macchi, L. (1995). Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental Psychology, 48A(1)*, 188-207.

Marks, D. F. and Clarkson, J. K. (1972). An explanation of conservatism in the bookbag-and-pokerchips situations. *Acta psychologica*, *36*, 145-160.

Noveck, I. A. (1997). Deductive competence need not be problematic: Some comments on Evan and Over's proposal concerning human rationality. *Current Psychology of Cognition, 17, 1*, 162-172.

Noveck, I. A., Lea, R. B., Davidson, G. M., & O'Brien, D. P. (1991). Human reasoning is both logical and pragmatic. *Intellectica, 11*, 81-109.

Piatelli Palmarini, M. (1993). L'illusione di sapere. Arnoldo Mondadori, Ed. S.p.A. Milano.

Politzer G. and Noveck, I. (1991). Are conjunction rule violations the result of conversational rule violaions. *Journal of Psycholinguistic Research, 20*, 83-103.

Schwarz, N., Strack, F., Hilton, D. and Naderer, G. (1991). Base rates, representativeness, and the logic of conversation, The contextual relevance of 'irrelevant' information. *Social Cognition, 9*, 67-83.

Slovic, P., Lichtenstein, S. and Fischhoff, B. (1988). Decision Making. In *Stevens's Handbook on of Experimental Psychology, Learning and Cognition*, Atkinson, R., Herrnstein, R. J., Lindzey, G. And Luce, R. D. Eds, Vol. 2, Chap.10, 673-737.

Sperber, D., Cara, F. and Giroto, V. (1995). Relevance theory explains the selection task. *Cognition, 57,* 31-95.

Sperber, D. and Wilson, D. (1986/1995). *Relevance: Communication and Cognition*, Blackwell, Oxford.

Teigen, K. H. (1974). Overestimation of subjective probabilities. *Scandinavian Journal of Psychology*, *15*, 56-62.

Teigen, K. H. (1983). Studies in subjective probability III: The unimportance of alternatives. *Scandinavian Journal of Psychology, 24*, 97-105.

Tversky, A. And Kahneman, D. (1982). Evidential impact of base rates. In: *Judgment under uncertainty: Heuristics and biases*, (pp. 153-160), eds. D. Kahneman, P. Slovic and A. Tversky, Cambridge University Press.

Van Wallendael, R.L. B. and Hastie, R. (1990). Tracing the foosteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory and Cognition, 18*, 240-250.

Von Winterfeldt, D. and Edwards, W, (1986). *Decision analysis and behavioral research*. Cambridge University Press.

Wells, G. L. and Harvey, J. H. (1978). Naive attributors' attributions and predictions: What is informative and when is an effect an effect? *Journal of Personality and Social Psycholog,y 36*, 483-490.

Zukier, H. and Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology, 47*, 349-360.

Table 1.

*Mean Judgments of the chances that the person described is a Math teacher (in percentages) and deviation from Bayesian predictions (N=120).*

| Portraits | $P_{Low}$ | $P_{High}$ | DMHL | BE-AE$_{Low}$ | BE-AE$_{High}$ | BE-AE$_{Mean}$ | $b_{Low}$ | $b_{High}$ | $b_{Mean}$ |
|---|---|---|---|---|---|---|---|---|---|
| *Standard Condition* | | | | | | | | | |
| Anne | 20.4 | 29.9 | 9.5 | 10.2 | 21.7 | 16.0 | 48 | 30 | 39 |
| Françoise | 38.2 | 44.2 | 6.0 | 23.0 | 25.1 | 24.1 | 21 | 19 | 20 |
| Jacques | 64.9 | 73.2 | 8.3 | 23.7 | 15.0 | 19.4 | 26 | 35 | 31 |
| Paul | 52.9 | 59.1 | 6.3 | 25.2 | 25.1 | 25.1 | 20 | 20 | 20 |
| Raphaël | 40.6 | 44.5 | 4.0 | 22.9 | 29.0 | 26.0 | 15 | 12 | 13 |
| *Mean* | 43.4 | 50.2 | 6.8 | 21.0 | 23.2 | 22.1 | 24 | 23 | 24 |
| *Suggested Complementarity  Condition* | | | | | | | | | |
| Anne | 19.1 | 32.1 | 13.0 | 8.6 | 17.9 | 13.3 | 60 | 42 | 51 |
| Françoise | 40.0 | 49.6 | 9.6 | 22.1 | 20.4 | 21.3 | 30 | 32 | 31 |
| Jacques | 60.3 | 74.3 | 14.1 | 18.5 | 8.3 | 13.4 | 43 | 63 | 53 |
| Paul | 46.0 | 57.0 | 11.0 | 17.0 | 21.1 | 19.1 | 39 | 34 | 37 |
| Raphaël | 37.5 | 45.9 | 8.4 | 16.3 | 25.9 | 21.1 | 34 | 24 | 29 |
| *Mean* | 40.6 | 51.8 | 11.2 | 16.5 | 18.7 | 17.6 | 40 | 37 | 39 |
| *Induced Complementarity Condition* | | | | | | | | | |
| Anne | 17.8 | 34.5 | 16.8 | 5.7 | 13.4 | 9.5 | 75 | 56 | 65 |
| Françoise | 37.5 | 53.3 | 15.8 | 16.7 | 17.1 | 16.9 | 49 | 48 | 48 |
| Jacques | 56.8 | 76.3 | 19.5 | 14.2 | 7.3 | 10.7 | 58 | 73 | 65 |
| Paul | 41.0 | 60.0 | 19.0 | 11.8 | 9.4 | 10.6 | 62 | 67 | 64 |
| Raphaël | 36.3 | 52.8 | 16.5 | 12.1 | 16.8 | 14.4 | 58 | 50 | 54 |
| *Mean* | 37.9 | 55.4 | 17.5 | 12.1 | 12.8 | 12.4 | 59 | 58 | 58 |

Notes. The Low and High subscripts refer to the experimental conditions which refer to base rates of mathematics teachers. Thus, $P_{Low}$ provides the mean probabilities that the person described is a mathematics teacher given the Low base rate condition. $P_{High}$ presents such mean probabilities in the High base rate condition. The DMHL is the ***D****ifference between the* ***M****eans of the* ***H****igh and* ***L****ow Base Rate. BE-AE* is the difference between the *Bayesian Estimate* and the *Actual Estimate* (see text). The index *b* varies from 0 (no indication of base rate utilization) to 100 (perfect Bayesian treatment of base rate information).

Table 2.

*Mean Judgments of the chances that the person described is a Math teacher (in percentages) and deviation from Bayes's prediction in the Suggested-Complementarity condition after having removed those (11) participants who violated the complementarity constraint (N=29).*

| Portraits | $P_{Low}$ | $P_{High}$ | DMHL | $BE\text{-}AE_{Low}$ | $BE\text{-}AE_{High}$ | $BE\text{-}AE_{Mean}$ | $b_{Low}$ | $b_{High}$ | $b_{Mean}$ |
|---|---|---|---|---|---|---|---|---|---|
| *Refined complementarity set* | | | | | | | | | |
| Anne | 20.8 | 37.5 | 16.7 | 8.4 | 16.2 | 12.3 | 67 | 51 | 59 |
| Françoise | 38.1 | 51.3 | 13.2 | 19.4 | 18.8 | 19.1 | 40 | 41 | 41 |
| Jacques | 58.5 | 75.6 | 17.2 | 16.1 | 7.9 | 12.0 | 52 | 69 | 60 |
| Paul | 41.9 | 61.6 | 19.6 | 13.0 | 14.2 | 13.6 | 60 | 58 | 59 |
| Raphaël | 34.6 | 49.7 | 15.1 | 14.7 | 20.5 | 17.6 | 51 | 42 | 47 |
| *Mean* | 38.8 | 55.1 | 16.4 | 14.3 | 15.5 | 14.9 | 53 | 52 | 53 |

Notes. See Table 1.

Table 3.

*Number and kind of violations of the complementarity constraint with respect to portraits in the Suggested-Complementarity condition.*

| | *Low Group* | | *High Group* | | *Subtotal* | | *Total* |
|---|---|---|---|---|---|---|---|
| *Portrait* | *Sub* | *Super* | *Sub* | *Super* | *Sub* | *Super* | |
| Anne | 4 | 0 | 3 | 1 | 7 | 1 | 8 |
| Françoise | 3 | 2 | 2 | 0 | 5 | 2 | 7 |
| Jacques | 4 | 1 | 2 | 1 | 6 | 2 | 8 |
| Paul | 2 | 2 | 2 | 0 | 4 | 2 | 6 |
| Raphaël | 1 | 1 | 3 | 0 | 4 | 1 | 5 |
| *Total* | 14 | 6 | 12 | 2 | 26 | 8 | 34 |

Notes. *Sub* and *Super* refer to subadditive and superadditive, respectively. Errors were committed by 11 subjects.

Table 4. *Mean judgments of the chances that the person described is a Math teacher (in percentages) and their deviation from Bayes's prediction, Experiment 2 (N=40).*

| Portraits | $p_{Low}$ | $p_{High}$ | DMHL | BE-AE$_{Low}$ | BE-AE$_{High}$ | BE-AE$_{Mean}$ | $b_{Low}$ | $b_{High}$ | $b_{Mean}$ |
|---|---|---|---|---|---|---|---|---|---|
| *Urn Condition* | | | | | | | | | |
| Anne | 19.5 | 35.00 | 15.5 | 7.37 | 11.67 | 9.52 | 68 | 57 | 62 |
| Françoise | 39 | 49.50 | 10.5 | 13.75 | 19.38 | 16.56 | 43 | 35 | 39 |
| Jacques | 58.25 | 74.75 | 16.5 | 12.45 | 8.17 | 10.31 | 57 | 67 | 62 |
| Paul | 44 | 60.25 | 16.25 | 16.25 | 11.92 | 14.08 | 50 | 58 | 54 |
| Raphaël | 37 | 51.75 | 14.75 | 11.06 | 16.72 | 13.89 | 57 | 47 | 52 |
| *Mean* | 39.55 | 54.25 | 14.70 | 12.18 | 13.57 | 12.87 | 55 | 52 | 53 |
| *Refined Urn condition* | | | | | | | | | |
| Anne | 19.5 | 35.79 | 16.29 | 6.96 | 10.88 | 8.92 | 70 | 60 | 65 |
| Françoise | 39 | 50.00 | 11.00 | 13.00 | 18.88 | 15.94 | 46 | 37 | 41 |
| Jacques | 58.25 | 75.79 | 17.54 | 11.00 | 7.13 | 9.07 | 61 | 71 | 66 |
| Paul | 44 | 62.89 | 18.89 | 14.89 | 9.27 | 12.08 | 56 | 67 | 62 |
| Raphaël | 37 | 52.11 | 15.11 | 10.39 | 16.37 | 13.38 | 59 | 48 | 54 |
| *Mean* | 39.55 | 55.32 | 15.77 | 11.25 | 12.51 | 11.88 | 58 | 56 | 57 |

Note. All instructions requested complementary estimates (exactly like those in the Suggested-complementarity condition of Experiment 1). The refined summary removes the single participant who violated the complementarity constraint. For other explanations, see Table 1.

Table 5.

*Mean Judgments of the chances that the person described is a Math teacher (in percentages) and deviation from Bayesian predictions, Experiment 3 (N=128).*

| Portraits | $p_{Low}$ | $p_{High}$ | DMHL | BE-AE$_{Low}$ | BE-AE$_{High}$ | BE-AE$_{Mean}$ | $b_{Low}$ | $b_{High}$ | $b_{Mean}$ |
|---|---|---|---|---|---|---|---|---|---|
| *Urn, Collective Condition* | | | | | | | | | |
| Anne | 19,1 | 37,8 | 18,8 | 6,8 | 14,3 | 10,5 | 73 | 57 | 65 |
| Françoise | 35,6 | 52,5 | 16,9 | 15,3 | 20,2 | 17,7 | 53 | 46 | 49 |
| Jacques | 57,5 | 76,3 | 18,8 | 14,3 | 9,5 | 11,9 | 57 | 66 | 62 |
| Paul | 40,9 | 60,3 | 19,4 | 15,3 | 14,9 | 15,1 | 56 | 56 | 56 |
| Raphaël | 35,6 | 53,8 | 18,1 | 13,9 | 17,9 | 15,9 | 57 | 50 | 54 |
| *Mean* | 37,8 | 56,1 | 18,4 | 13,1 | 15,4 | 14,2 | 58 | 54 | 56 |
| *Urn, Individual Condition* | | | | | | | | | |
| Anne | 20,0 | 36,9 | 16,9 | 7,4 | 12,5 | 10,0 | 69 | 57 | 63 |
| Françoise | 36,3 | 50,6 | 14,4 | 13,4 | 15,1 | 14,2 | 52 | 49 | 50 |
| Jacques | 58,8 | 76,9 | 18,1 | 13,0 | 6,0 | 9,5 | 58 | 75 | 67 |
| Paul | 43,8 | 61,6 | 17,8 | 12,6 | 9,1 | 10,8 | 59 | 66 | 62 |
| Raphaël | 37,5 | 52,8 | 15,3 | 13,0 | 19,4 | 16,2 | 54 | 44 | 49 |
| *Mean* | 39,3 | 55,8 | 16,5 | 11,9 | 12,4 | 12,2 | 58 | 57 | 58 |
| *no-Urn*, Collective Condition* | | | | | | | | | |
| Anne | 19,8 | 33,3 | 13,6 | 9,0 | 19,2 | 14,1 | 60 | 41 | 51 |
| Françoise | 36,9 | 46,1 | 9,1 | 21,7 | 26,5 | 24,1 | 30 | 26 | 28 |
| Jacques | 59,7 | 74,1 | 14,4 | 22,4 | 11,2 | 16,8 | 39 | 56 | 48 |
| Paul | 48,3 | 57,8 | 9,6 | 26,4 | 22,5 | 24,5 | 27 | 30 | 28 |
| Raphaël | 40,4 | 53,8 | 13,4 | 20,5 | 21,4 | 20,9 | 39 | 39 | 39 |
| *Mean* | 41,0 | 53,0 | 12,0 | 20,0 | 20,1 | 20,1 | 37 | 37 | 37 |
| *no-Urn, Individual Condition* | | | | | | | | | |
| Anne | 20,3 | 31,3 | 11,0 | 9,4 | 19,8 | 14,6 | 54 | 36 | 45 |
| Françoise | 37,7 | 45,3 | 7,6 | 21,6 | 23,2 | 22,4 | 26 | 25 | 25 |
| Jacques | 61,1 | 73,2 | 12,1 | 15,8 | 7,6 | 11,7 | 43 | 61 | 52 |
| Paul | 49,3 | 59,2 | 9,9 | 22,3 | 19,8 | 21,0 | 31 | 33 | 32 |
| Raphaël | 36,8 | 49,2 | 12,4 | 13,1 | 21,2 | 17,2 | 49 | 37 | 43 |
| *Mean* | 41,0 | 51,6 | 10,6 | 16,4 | 18,3 | 17,4 | 39 | 37 | 38 |

Notes. *The no-Urn condition is identical to the Suggested-complementarity condition of Experiment 1. All conditions requested two estimates. For other explanations, see Table 1.